

CORDEX Writer (v1)

Grigory Nikulin and Michael Kolax

Rosby Centre, SMHI

February 20, 2013

1. Introduction

A set of scripts is designed to produce the CORDEX CORE and Tier 1 output in netcdf format. All scripts are bash-shell scripts and use the CDO and NCO packages. This document gives a short practical introduction to the scripts, describing how to configure, modify and run them to produce CORDEX compliant netcdf files. A basic idea behind the scripts is to use all information already available in raw output from a regional climate model - RCA4 (Rosby Centre, SMHI). Since raw output from different RCMs is different in many aspects some modifications are needed for a specific RCM. The scripts present one of many ways how CORDEX compliant files can be produced and their parts can be used for including in post-processing of RCM output.

The CORDEX archive specifications can be found at cordex.dmi.dk

http://cordex.dmi.dk/joomla/images/CORDEX/cordex_archive_specifications.pdf

http://cordex.dmi.dk/joomla/images/CORDEX/CORDEX_variables_requirement_table.pdf

Limited support can be provided (contact - grigory.nikulin@smhi.se).

2. Overall structure of the CORDEX writer

The CORDEX writer package includes a few scripts to produce daily, monthly and seasonal mean output.

Root directory includes three main scripts

- `cordex_writer_day_v1.sh` - daily script (wrapper)
- `cordex_day2mon_v1.sh` - monthly script
- `cordex_mon2sem_v1.sh` - seasonal script

and following sub-directories

- `writer_meta` configuration and subroutine files
- `cordex_tables` the CORDEX MIP tables
- `input files` a few example grib files used as an input to the CORDEX writer

- `output_files` netcdf files produced by the CORDEX writer.

3. Required components to install

- CDO (<https://code.zmaw.de/projects/cdo>). CDO version 1.5.4 is used.
- NCO (<http://nco.sourceforge.net/>). NCO version 4.0.6 is used.
- BASH version 3.2.25 is used.
- CORDEX MIP tables

Although all tables are provided with the CORDEX writer, it is recommended to download the latest CORDEX tables and place them in `cordex_tables` sub-directory. Note that only one table - `CORDEX_day` is used in the CORDEX writer and all differences in the monthly and seasonal output compared to the daily one are directly fixed in respective scripts. http://www2-pcmdi.llnl.gov/cmor/tables/copy4_of_cmip5-tables/

4. Configuration files and environment variables

- Configuration file with simulation information

`run_info_cordex.sh` is located in `writer_meta` sub directory and provides necessary information to create the CORDEX file name structure and global attributes. Each simulation has a short ID. For example “201116” corresponds to the ERA-Interim downscaling by SMHI-RCA4 and used in all scripts.

- Configuration file with variable information

All variable information (standard and long names, units and other attributes) is taken from the CORDEX MIP tables. However, some additional information is needed for processing grib files (grib codes) or for processing netcdf files with other variable names than in the CORDEX specifications. Such information is provided in `var_info_cordex.sh` and used only once to rename variables accordingly the CORDEX specifications. In case of RCA4 this file contains internal grib codes used in RCA4.

- Environmental variable - **CORDEX_WRITER_PATH** should be set to the root directory with the CORDEX writer scripts.

5. General comments

Originally, the daily script was developed to post-process raw grib files from a regional climate model - RCA4 (Rossby Centre, SMHI). The CORDEX writer version of the daily script was decoupled from the RCA4 environment but still few requirements/limitations left.

- Since the RCA4 output is in grib format input files to the daily script must be in grib or netcdf format. The input files should be already post-processed (daily averaging/accumulation, 5-year periods etc.) and do have the time coordinate. Such first step in post-processing is completely dependent on RCM-Institute local computer environment (how output is saved) and not included in the CORDEX writer scripts.
- The RCA4 model uses the rotated coordinate system. The daily script converts the rotated coordinates to more usual 2D geographical ones and adds the geographical coordinates to output files. Information about the rotated coordinates is taken directly from grib files and post-processed completely automatically. There is no need to create a separate file with the 2D geographical coordinates. However, for other coordinates systems than the rotated one (Lambert conformal, Mercator etc.) this approach does not work and the 2D geographical coordinates should be added by other means (copying from a separate file for example).

6. Daily output

`cordex_writer_day_v1.sh` is a wrapper script which calls the main daily script `cordex_raw2day_v1.sh` located in `writer_meta` sub-directory. As a first testing step it is recommended to run the daily script as it is. If all requirement components are installed the script post-processes five input files (tas, pr, psl, rsds and psl for 1980) saving output files to `/output_files/day`.

Comments to `cordex_raw2day_v1.sh`

- **DATES in FILE NAME: YYYYMMDD**

Dates in output files names are directly taken from the first and last time step in input files, no modification is needed.

- **CONVERT to netcdf and fix TIME**

If input files are already in netcdf format “-f nc” option should be removed. Leaving this option does not influence output files but may increase post-processing time.

The missing values are set to 1.e20 but should already be defined in input files.

Time units are set to “days since 1949-12-01,00:00”.

Calendar is set accordingly to variable \$calendar ('standard', '365days', '360days' or 'proleptic') provided in the configuration file `run_info_cordex.sh`.

Time (hours) is set to 12:00 for all time steps.

- **2D LONGITUDE and LATITUDE**

2D longitude and latitude coordinates are calculated and added to output files.

This works only for RCMs with the rotated coordinates and even in this case functionality might depend on how coordinate information is presented in input files.

For RCM with other coordinate system than the rotated one the 2D real geographical coordinates should be added by other means, coping from a pre-created file, for example.

- **TIME ATTRIBUTES**

Time attributes are created or overwritten. No modifications are needed.

- **TIME BOUNDS**

Time bounds (time_bnds) are added. No modifications are needed.

- **VARIABLE ATTRIBUTES**

“Old” variable names defined in `var_info_cordex.sh` are changed to the CORDEX variable names. This might be removed if input files are in netcdf format and already have the CORDEX variable names.

Variable attributes are created or overwritten. Additional attributes can be added (comment for example, etc.).

- **VERTICAL COORDINATE**

This part is strongly dependent on how RCM output is configured for a given RCM. For some variables, which have no vertical coordinate, like precipitation (pr), sea level pressure (psl) and short wave radiation (rsds) in the given example script CDO creates a vertical coordinate “height” or “alt”. This feature is only related the RCA4 output configuration and can be pretty different in other RCMs.

The unnecessary vertical dimension and corresponding variable can be easily removed, in case of “height”, by next two lines:

```
ncwa -O -a height -h $file_out $file_out
```

```
ncks -O -x -v height -h $file_out $file_out
```

For all variables with vertical dimensions “height” or “plev” corresponding variable is created.

- **PLEV_BNDS for CLOUDS in LAYERS**

Pressure level bounds for variables “cll”, “clm” and “clh” are created. No modifications are needed.

- **GLOBAL ATTRIBUTES**

Global attributes are created or overwritten. The most attributes are taken from the configuration file - `run_info_cordex.sh`.

Attributes “contact”, “institution”, “institute_id” and “references” should be changed in `run_info_cordex.sh`.

Two attributes “rossby_comment” and “rossby_run_id” can be removed or renamed.

The global attribute “tracking_id” (commented) provides a random unique number generated by “`uuid -v 4`” command. This attribute is not defined in the CORDEX archive specification but recommended by analogy to the CMIP5 output.

- **REMOVE GLOBAL ATTRIBUTES**

Unnecessary global attribute are deleted.

7. Monthly output

`cordex_day2mon_v1.sh` takes the daily output produced by `cordex_writer_day_v1.sh` (`cordex_raw2day_v1.sh`). Generally, it can be any daily netcdf files produced by any means but accordingly to the CORDEX specification for the Tier1.

Accordingly to the CORDEX archive design “For monthly and seasonal data, 10 years with one file for each averaging period. Files should start and end at “round numbers”. For instance, a simulation 1953-1979 should have monthly files corresponding to 1953-1960, 1961-1970 and 1971-1979.”

Exact definition for output periods can be done by setting up variables `$fy_in`, `$ly_in` and `$sy_in` and few most common examples – ERA-Interim 1980-2010, Historical 1951-2005 and Scenario 2006-2100 are provided.

For RCMs with the rotated coordinates the monthly script can be run as it is and no modification are needed.

- **HORIZONTAL COORDINATES**

For RCMs with other coordinate systems a few modification should be included. Monthly averaging is efficiently done by CDO but CDO cannot safely post-processed netcdf files with two coordinate systems provided, a part of this information (one coordinate system) is lost. Additionally, dimension names might be changed by CDO. To fix the problem the dimension names are renamed and the lost coordinate information is simply copied from one of input files to all output files. This part is dependent on a coordinate system used.

8. Seasonal output

Seasonal script `cordex_mon2sem_v1.sh` takes input from the monthly script and is pretty similar to the monthly one.

No definition of time periods for output files are needed. This is defined automatically but monthly input must follow the CORDEX specifications.

- **HORIZONTAL COORDINATES**

Modifications are needed for RCMs with other coordinate system than the rotated one. See comments for the monthly script.